MEASURES OF ASSOCIATION FOR CROSS CLASSIFICATIONS*

LEO A. GOODMAN AND WILLIAM H. KRUSKAL
University of Chicago

CONTENTS

* This paper is partly an outgrowth of work sponsored by the Army, Navy, and Air Force through the Joint Services Advisory Committee for Research Groups in Applied Mathematics and Statistics by Contract No. N6ori-02035. We are indebted for helpful comments and criticisms to: Otis D. Duncan (University of Chicago), Churchill Eisenhart (National Bureau of Standards), Maurice G. Kendall

When populations are cross-classified with respect to two or more classifications or polytomies, questions often arise about the degree of association existing between the several polytomies. Most of the traditional measures or indices of association are based upon the standard chi-square statistic or on an assumption of underlying joint normality. In this paper a number of alternative measures are considered, almost all based upon a probabilistic model for activity to which the cross-classification may typically lead. Only the case in which the population is completely known is considered, so no question of sampling or measurement error appears. We hope, however, to publish before long some approximate distributions for sample estimators of the measures we propose, and approximate tests of hypotheses. Our major theme is that the measures of association used by an empirical investigator should not be blindly chosen because of tradition and convention only, although these factors may properly be given some weight, but should be constructed in a manner having operational meaning within the context of the particular problem.

1. INTRODUCTION

MANY studies, particularly in the social sciences, deal with populations of individuals which are thought of as cross-classified by two or more polytomies. For example, the adult individuals living in New York City may be classified as to

Borough:	5 classes
Newspaper most often read:	perhaps 6 classes
Television set in home or not:	2 classes
Level of formal education:	perhaps 5 classes
Age:	perhaps 10 classes

For simplicity we deal largely with the case of two polytomies, although many of our remarks may be extended to a greater number. The double polytomy is the most common, no doubt because of the ease with which it can be tabulated and displayed on the printed page. Most of our remarks suppose the population completely known in regard to the classifications, and indeed this seems to be the way to begin in the construction of rational measures of association. After agreement has been reached on the utility of a measure for a known population, then

(London School of Economics and Political Science), Frederick Mosteller (Harvard University), I. Richard Savage (National Bureau of Standards), Alan Stuart (London School of Economics and Political Science), Louis L. Thurstone (University of North Carolina), John W. Tukey (Princeton (University), W. Allen Wallis (University of Chicago), and E. J. Williams (Commonwealth Scientific and Industrial Research Organization, Australia). Part of Mr. Goodman's work on this paper was carried out at the Statistical Laboratory of the University of Cambridge under a Fulbright Award and a Social Science Research Council Fellowship. The authors were led to work on the problems of this paper as a result of conversations with Louis L. Thurstone and Bernard R. Berelson.

one should consider the sampling problems associated with estimation and tests about this population parameter.

A double polytomy may be represented by a table of the following kind:[1]

A	B				
	B_1	B_2	$\cdots$	B_β	Total
A_1	ρ_{11}	ρ_{12}	$\cdots$	$\rho_{1\beta}$	$\rho_{1\cdot}$
A_2	ρ_{21}	ρ_{22}	$\cdots$	$\rho_{2\beta}$	$\rho_{2\cdot}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
A_α	$\rho_{\alpha 1}$	$\rho_{\alpha 2}$	$\cdots$	$\rho_{\alpha\beta}$	$\rho_{\alpha\cdot}$
Total	$\rho_{\cdot 1}$	$\rho_{\cdot 2}$	$\cdots$	$\rho_{\cdot\beta}$	1

where

Classification A divides the population into the α classes $A_1, A_2, \cdots, A_\alpha$.

Classification B divides the population into the β classes $B_1, B_2, \cdots, B_\beta$.

The proportion of the population that is classified as both A_a and B_b is ρ_{ab}.

The marginal proportions will be denoted by

$\rho_{a\cdot}$ = the proportion of the population classified as A_a.

$\rho_{\cdot b}$ = the proportion of the population classified as B_b.

If the use to which a measure of association were to be put could be precisely stated, there would be little difficulty in defining an appropriate measure. For example, using the above cross-classification of the New York City population, a television service company might wish to

[1] Tables of this kind are frequently called *contingency tables*. We shall not use this term because of its connotation of a specific sampling scheme when the population is not known and one infers on the basis of a sample.

place a single newspaper advertisement which would be read by as many prospective customers as possible. Then the important information from the table of newspaper-most-often-read vs. television-set-in-home-or-not would be: which newspaper is most often read among those with television sets? And a reasonable measure of association would simply be the proportion of those with television sets who read this newspaper.

It is rarely the case, however, that the purpose of an investigation can be so specifically stated. More typically an investigation is exploratory or has a multiplicity of goals. Sometimes a measure of association is desired simply so that a large mass of data may be summarized compactly.

The basic theme of this paper is that, even though a single precise goal for an investigation cannot be specified, it is still possible and desirable to choose a measure of association which has contextual meaning, instead of using as a matter of course one of the traditional measures. In order to choose a measure of association which has meaning we propose the construction of probabilistic models of predictive activity, the particular model to be chosen in the light of the particular investigation at hand. The measure of association will then be a probability, or perhaps some simple function of probabilities, within such a model. Such is our general contention; most of the remainder of this paper is concerned with its exemplification in particular instances.

We wish to emphasize that the specific measures of association described here are *not* presented as factotum or universal measures. Rather, they are suggested as reasonable for use in appropriate circumstances only, and even in those circumstances other measures may and should be considered and investigated.

A good deal of attention has been paid in the literature to the special case of two dichotomies. We are more interested here in measures of association suitable for use with any numbers of classes in the polytomies or classifications.

2. FOUR PRELIMINARY CONSIDERATIONS

Four distinctions or cautionary remarks should be made early in any discussion of measures of association.

2.1. *Continua*

We may or may not wish to think of a polytomy as arising from an underlying continuum. For example, age may for convenience be divided into ten classifications, but it clearly does arise from an underlying continuum; however, newspaper-most-often-read would scarcely be so construed. If a polytomy does arise from an underlying continuum one may or may not wish to assume that the population has some specific kind of distribution with respect to it.

In those cases in which all the polytomies of a study arise jointly from a multivariate normal distribution on an underlying continuum, one would naturally turn to measures of association based on the correlation coefficients. These in turn might well be estimated from a sample by the tetrachoric correlation coefficient method or a generalization of it. In some cases one polytomy may arise from a continuum and the other not. An interesting discussion of this case for two dichotomies was given in 1915 by Greenwood and Yule ([3], Section 3). We do not discuss either of these cases in this paper, but restrict ourselves to situations in which there are no relevant underlying continua.

The desirability of assuming an underlying joint continuum was one of the issues of a heated debate forty years ago between Yule [15] on the one hand and K. Pearson and Heron [9] on the other. Yule's position was that very frequently it is misleading and artificial to assume underlying continua; Pearson and Heron argued that almost always such an assumption is both justified and fruitful.

2.2. *Order*

There may or may not be an underlying order between the classifications of a polytomy. For example "level of formal education" admits an obvious ordering; but borough of residence would not usually be thought of in an ordered way. If there is an ordering, it may or may not be relevant to the investigation. Sometimes an ordering may be important but not its direction. If there is an underlying one-dimensional continuum, it establishes an ordering.

When there is no natural or relevant ordering of the classes of a polytomy, one may reasonably ask that a measure of association not depend on the particular order in which the classes are tabulated.

2.3. *Symmetry*

It may or may not be that one looks at two polytomies symmetrically. When we are sure a priori that a causal relationship (if it exists) runs in one direction but not the other, then our viewpoint will be asymmetric. This will also happen if one plans to *use* the results of the experiment in one direction only. On the other hand, there is often no reason to give one polytomy precedence over another.

2.4. *Manner of Formation of the Classes*

Decisions about the definitions of the classes of a polytomy, or changes from a finer to a coarser classification (or vice-versa), can affect all the measures of association of which we know. For example, suppose we begin with the 4×4 table

0	.25	0	0
.25	0	0	0
0	0	0	.25
0	0	.25	0

and combine neighboring pairs of classes. We obtain

.5	0
0	.5

which might greatly change a measure of association. Or we might combine the three bottom rows and the three right-hand columns. This gives

0	.25
.25	.5

which presents quite a different intuitive degree of association. By other poolings one can obtain other 2×2 tables.

Although this example is extreme, similar changes can be made in the character of almost any cross-classification table. Related examples are discussed by Yule [15].

At first this consideration might seem to vitiate any reasonable discussion of measures of association. We feel, however, that it is in fact desirable that a measure of association reflect the classes *as defined for* *the data*. Thus one should not speak, for example, of association between income level and level of formal education without specifying particular class definitions. Of course, in many cases association—however measured—would not be much affected by any reasonable redefinition of the classes, and then the above finicky form of statement can be simplified. That the definition of the classes can affect the degree of association naturally means that careful attention should be given to the class definitions in the light of the expected uses of the final conclusions.

3. CONVENTIONS

It is conventional, and often convenient, to set up a measure of association so that either

(*i*) It takes values between -1 and $+1$ inclusive, is -1 or $+1$ in case of "complete association," and is zero in the case of independence.
(*ii*) It takes values between 0 and $+1$ inclusive, is $+1$ in the case of "complete association," and is zero in the case of independence.

Convention (*i*) is appropriate when the association is thought of as signed (e.g., association between income and dollars spent is positive, between income and per cent of income spent is negative). Convention (*ii*) is appropriate when no such sign considerations exist, as when there is no natural order.

"Complete association," as we shall see, is somewhat ambiguous. "Independence," on the other hand, has its usual meaning, that is

$$\rho_{ab} = \rho_{a\cdot}\rho_{\cdot b} \; (a = 1, \cdots, \alpha; b = 1, \cdots, \beta). \tag{1}$$

Conventions like these have seemed important to some authors, but we believe they diminish in importance as the meaningfulness of the measure of association increases. One real danger connected with such conventions is that the investigator may carry over size preconceptions based upon experience with completely different measures subject to the same conventions. For example, some elementary statistics textbooks warn that a population correlation coefficient less than about .5 in absolute value may have little practical significance, in the sense that then the conditional variance is not much less than the marginal variance. Research workers in various fields thus tend to develop rather strong feelings that population correlation coefficients less than, say, .5, have little substantive importance. The same feelings might be carried over, without justification, to all other measures of association so defined as to lie between $+1$ and -1.

It should also be mentioned that once one has a measure of association satisfying one of the above conventions, then an infinite number of others also satisfying the same convention can be obtained—for example, by raising to a power and adjusting the sign if necessary.

4. TRADITIONAL MEASURES

Excellent accounts of these may be found in [16], Chaps. 2 and 3, and [7], Chap. 13. Many of these stem from the standard chi-square statistic upon which a test of independence is usually based. If a finite population has ν members and we set $\nu_{ab}=\nu\rho_{ab}$, $\nu_{a\cdot}=\nu\rho_{a\cdot}$, $\nu_{\cdot b}=\nu\rho_{\cdot b}$, etc., the chi-square statistic in the case of two classifications is

$$\chi^2 = \sum_a \sum_b \frac{(\nu_{ab} - \nu_{a\cdot}\nu_{\cdot b}/\nu)^2}{\nu_{a\cdot}\nu_{\cdot b}/\nu} = \nu \sum_a \sum_b \frac{(\rho_{ab} - \rho_{a\cdot}\rho_{\cdot b})^2}{\rho_{a\cdot}\rho_{\cdot b}} \tag{2}$$

$$= \nu \sum_a \sum_b \frac{\rho_{ab}^2}{\rho_{a\cdot}\rho_{\cdot b}} - \nu.$$

A great deal of attention has been given to the case $\alpha=\beta=2$. For this special case Yule has defined the following coefficient of association:

$$Q = \frac{\nu_{11}\nu_{22} - \nu_{12}\nu_{21}}{\nu_{11}\nu_{22} + \nu_{12}\nu_{21}} \tag{3}$$

whose numerator squared is essentially the same as that of a convenient and popular form for χ^2 in the 2×2 case. Another coefficient suggested by Yule for the 2×2 case is

$$Y = \frac{\sqrt{\nu_{11}\nu_{22}} - \sqrt{\nu_{12}\nu_{21}}}{\sqrt{\nu_{11}\nu_{22}} + \sqrt{\nu_{12}\nu_{21}}}\,. \tag{4}$$

A coefficient often used for the general $\alpha\times\beta$ case is simply χ^2/ν, often called the mean square contingency and denoted by ϕ^2. A variation of this, suggested by Karl Pearson, is

$$C = \sqrt{\frac{[\chi^2/\nu]}{1 + \chi^2/\nu}} \tag{5}$$

which has been called the coefficient of contingency, or the coefficient of mean square contingency. Another variation, proposed by Tschuprow, is

$$T = \sqrt{[\chi^2/\nu]/(\alpha - 1)(\beta - 1)}. \tag{6}$$

The last two suggestions, according to Kendall [7], were made in attempts to norm χ^2 so that it might lie between 0 and 1 and take the extreme values under independence and "complete association." Cramér ([1], p. 282) suggests the following variant:

$$[\chi^2/\nu]/\text{Min } (\alpha - 1, \beta - 1) \tag{7}$$

which gives a better norming than does C or T since it lies between 0 and 1 and actually attains both end points appropriately. Cramér's suggestion does not seem to be well known by workers using this general kind of index.

The fact that an excellent test of independence may be based on χ^2 does not at all mean that χ^2, or some simple function of it, is an appropriate *measure* of degree of association. A discussion of this point is presented by R. A. Fisher ([2], Section 21). We have been unable to find any convincing published defense of χ^2-like statistics as measures of association.

One difficulty with the use of the traditional measures, or of any measures that are not given operational interpretation, is that it is difficult to compare meaningfully their values for two cross-classifications. Suppose that C turns out to be .56 and .24 respectively in two cross-classification tables. One wants to be able to say that there is higher association in the first table than the second, but investigators sometimes restrain themselves, with commendable caution, from making such a comparison. Their restraint may stem in part from the noninterpretability of C. (Of course, when samples are small they may also be restrained by inadequate knowledge of sampling fluctuation.)

One class of measures that will not be discussed here is characterized by the assignment of numerical scores to the classes, followed by the use of the correlation coefficient on these scores. A recent article on such measures is by E. J. Williams [12]. It contains references leading back to earlier literature. We feel that the use of arbitrary scores to motivate measures is infrequently appropriate, but it should be pointed out that measures not motivated by the correlation of scores can often be thought of from the score viewpoint.

5. MEASURES BASED ON OPTIMAL PREDICTION

5.1. *Asymmetrical Optimal Prediction. A Particular Model of Activity*

Let us consider first a probabilistic model which might be useful in a situation of the following kind:

(*i*) Two polytomies, A and B.
(*ii*) No relevant underlying continua.
(*iii*) No natural ordering of interest.
(*iv*) Asymmetry holds: The A classification precedes the B classification chronologically, causally, or otherwise.

An example of such a situation might be a study of the association between college attended (A) and kind of adult occupation (B). Our model of activity is the following: An individual is chosen at random from the population and we are asked to guess his B-class as well as we can, either

1. Given no further information, or
2. Given his A class.

Clearly we can do no worse in case 2 than in case 1. Represent by $\rho_{\cdot m}$ the largest marginal proportion among the B classes and by ρ_{am} the largest proportion in the ath row of the cross-classification table—that is

$$(8) \qquad \rho_{\cdot m} = \operatorname*{Max}_b \rho_{\cdot b}, \quad \rho_{am} = \operatorname*{Max}_b \rho_{ab} \cdot$$

Then in case 1 we are best off guessing that B_b for which $\rho_{\cdot b} = \rho_{\cdot m}$—that is, guessing that B class which has the largest marginal proportion—and our probability of error is $1 - \rho_{\cdot m}$. In case 2 we are best off guessing that B_b for which $\rho_{ab} = \rho_{am}$ (letting A_a be the given A class)—that is, guessing that B class that has the largest proportion in the observed A class—and our probability of error is[2] $1 - \sum_a \rho_{am}$.

Then we propose as a measure of association (following Guttman [4])

$$(9) \qquad \lambda_b = \frac{(\text{Prob. of error in case 1}) - (\text{Prob. of error in case 2})}{(\text{Prob. of error in case 1})}$$

$$= \frac{\sum_a \rho_{am} - \rho_{\cdot m}}{1 - \rho_{\cdot m}},$$

which is the relative decrease in probability of error in guessing B_b as between A_a unknown and A_a known. To put this another way, λ_b gives the proportion of errors that can be eliminated by taking account of knowledge of the A classifications of individuals.

Some important properties of λ_b follow:

[2] It may be that in case 1 there is more than one b for which $\rho_{\cdot b} = \rho_{\cdot m}$. Then any method of choosing which of these b's to guess—including flipping an appropriately multi-sided die—gives rise to the same probability of error, $1 - \rho_{\cdot m}$. A similar comment applies to case 2.

(*i*) λ_b is indeterminate if and only if the population lies in one column, that is, lies in one B class.
(*ii*) Otherwise the value of λ_b is between 0 and 1 inclusive.
(*iii*) λ_b is 0 if and only if knowledge of the A classification is of no help in predicting the B classification, i.e., if there exists a b_0 such that $\rho_{ab_0} = \rho_{am}$ for all a.
(*iv*) λ_b is 1 if and only if knowledge of an individual's A class completely specifies his B class, i.e., if each row of the cross-classification table contains at most one nonzero ρ_{ab}.
(*v*) In the case of statistical independence λ_b, when determinate, is zero. The converse need not hold: λ_b may be zero without statistical independence holding.
(*vi*) λ_b is unchanged by permutation of rows or columns.

That λ_b may be zero without statistical independence holding may be considered by some as a disadvantage of this measure. We feel, however, that this is not the case, for λ_b is constructed specifically to measure association in a restricted but definite sense, namely the predictive interpretation given. If there is no association in that sense, even though there is association in other senses, one would want λ_b to be zero. Moreover, all the measures of association of which we know are subject to this kind of criticism in one form or another, and indeed it seems inevitable. To obtain a measure of association one must sharpen the definition of association, and this means that of the many vague intuitive notions of the concept some must be dropped.

We may similarly define

$$\lambda_a = \frac{\sum_b \rho_{mb} - \rho_{m\cdot}}{1 - \rho_{m\cdot}}, \tag{10}$$

where

$$\begin{aligned} \rho_{m\cdot} &= \operatorname*{Max}_a \rho_{a\cdot} \\ \rho_{mb} &= \operatorname*{Max}_a \rho_{ab}\,. \end{aligned} \tag{11}$$

Thus λ_a is the relative decrease in probability of error in guessing A_a as between B_b unknown and known.

So far as we know, λ_a and λ_b were first suggested by Guttman ([4], Part I, 4), and our development of them is very similar to his.

5.2. *Symmetrical Optimal Prediction. Another Model of Activity*

In many cases the situation is symmetrical, and one may alter the model of activity as follows: an individual is chosen at random from the population and we are asked to guess his A class half the time (at random) and his B class half the time (at random) either given:

1. No further information, or
2. The class of the individual other than the one being guessed; that is the individual's A_a when we guess B_b and vice versa.

In case 1 the probability of error is $1-\frac{1}{2}(\rho_{\cdot m}+\rho_{m\cdot})$, and in case 2 the probability of error is $1-\frac{1}{2}(\sum_a \rho_{am}+\sum_b \rho_{mb})$. Hence we may consider the relative decrease in probability of error as we go from case 1 to case 2, and define the coefficient

$$\lambda = \frac{\frac{1}{2}[\sum_a \rho_{am} + \sum_b \rho_{mb} - \rho_{\cdot m} - \rho_{m\cdot}]}{1 - \frac{1}{2}(\rho_{\cdot m} + \rho_{m\cdot})}. \tag{12}$$

Some properties of λ follow:

(i) λ is determinate except when the entire population lies in a single cell of the table.

(ii) Otherwise the value of λ is between 0 and 1 inclusive.

(iii) λ is 1 if and only if all the population is concentrated in cells no two of which are in the same row or column.

(iv) λ is 0 in the case of statistical independence, but the converse need not hold.

(v) λ is unchanged by permutations of rows or columns.

(vi) λ lies between λ_a and λ_b inclusive.

The computation of λ_a, λ_b, or λ is extremely simple. Usually one is given the population, not in terms of the ρ_{ab}'s but rather in terms of the numbers of individuals in each cell. Let ν be the total number of individuals in the population, $\nu_{ab}=\nu\rho_{ab}$, $\nu_{am}=\nu\rho_{am}$, $\nu_{mb}=\nu\rho_{mb}$, and so on. Then

$$\lambda_b = \frac{\sum_a \nu_{am} - \nu_{\cdot m}}{\nu - \nu_{\cdot m}}, \tag{13}$$

$$\lambda_a = \frac{\sum_b \nu_{mb} - \nu_{m\cdot}}{\nu - \nu_{m\cdot}}, \tag{14}$$

$$\lambda = \frac{\sum_a \nu_{am} + \sum_b \nu_{mb} - \nu_{\cdot m} - \nu_{m\cdot}}{2\nu - (\nu_{\cdot m} + \nu_{m\cdot})}. \tag{15}$$

5.3. *An example*

The following table is taken from reference [7], p. 300, and originally was given by Ammon in "Zur Anthropologie der Badener." It deals with hair and eye color of males. The table is given in terms of the ν_{ab}'s. A_1, A_2, A_3 are respectively Blue, Grey or Green, Brown; B_1, B_2, B_3, B_4 are respectively Fair, Brown, Black, Red.

Eye Color Group	Hair Color Group				
	B_1	B_2	B_3	B_4	$\nu_{a\cdot}$
A_1	1768	807	189	47	2811
A_2	946	1387	746	53	3132
A_3	115	438	288	16	857
$\nu_{\cdot b}$	2829	2632	1223	116	$\nu=6800$

We have:

$$\nu_{1m} = 1768 \qquad \nu_{m1} = 1768$$

$$\nu_{2m} = 1387 \qquad \nu_{m2} = 1387$$

$$\nu_{3m} = 438 \qquad \nu_{m3} = 746$$

$$\nu_{m4} = 53$$

$$\nu_{\cdot m} = 2829 \qquad \nu_{m\cdot} = 3132$$

$$\lambda_a = \frac{3,954 - 3,132}{6,800 - 3,132} = \frac{822}{3,668} = .2241$$

$$\lambda_b = \frac{3,593 - 2,829}{6,800 - 2,829} = \frac{764}{3,971} = .1924$$

$$\lambda = \frac{822 + 764}{3,668 + 3,971} = \frac{1,586}{7,639} = .2076.$$

(Quotients are given to four places.) The traditional measures of association have the following values: $\chi^2/\nu = .1581$, $C = .3695$, $T = .2541$, Cramér's measure $= .2812$.

This example appears as an illustration of the usual approach to measures of association in [7], a standard statistical reference work. It is not hard to think of interpretations or variations in which one of the λ coefficients would be appropriate. For example, one might be studying the efficacy of an identification scheme for males in which hair color was given but not eye color. Another example might be in connection with a study of popular beliefs about the relationship between hair color and eye color.

5.4. *Weighting Columns or Rows*

In some cases, particularly when comparisons between different populations are important, the measures λ_a, λ_b, or λ may not be suitable, since they depend essentially on the marginal frequencies. To put this in terms of the model of activity: in some cases we do not want to think of choosing an individual from the actual population *at hand* in a random way, but rather from some other population which is related to the actual population in terms of conditional frequencies.

This point is stressed by Yule in reference [15] and is illustrated by the kind of medical example[3] given there. Suppose that we are concerned with the effects of a medical treatment on persons contracting an often fatal disease. Very large samples from two different hospitals are available, giving the following ρ_{ab} tables:

	Hospital I			Hospital II		
	Lived	Died	Total	Lived	Died	Total
Treated	.84	.04	.88	.42	.02	.44
Not treated	.03	.09	.12	.14	.42	.56
Total	.87	.13	1.00	.56	.44	1.00

Here the A classes are Treated or Not-treated, and the B classes Lived or Died. The given numbers are ρ's and marginal ρ's.

We are interested in the association between treatment and life, and might conclude that λ_b would be an appropriate measure of this. We find

$$\lambda_b \text{ for Hospital I} = \frac{.93 - .87}{.13} = .462$$

$$\lambda_b \text{ for Hospital II} = \frac{.84 - .56}{.44} = .636.$$

[3] We do not wish to suggest by this example that λ_b is necessarily appropriate as a measure of association between treatment and cure. A very interesting discussion of this medical case has been given by Greenwood and Yule [3] who bring out many difficulties and suggest various viewpoints. Another interesting paper on the medical 2×2 table is that of Youden [14].

Yet the *conditional* probabilities of life, given treatment (nontreatment), are exactly the same for both hospitals, namely .955 (.250). The reason that the conditional probabilities are the same while the λ_b values are different is, of course, that the two hospitals treated very different proportions of their patients. And the proportions treated were probably determined by factors having nothing to do with 'inherent' association between treatment and cure.

It may seem reasonable in such a case as this to replace our model of activity by one in which an individual is drawn from the population so that the probability of his being in any given A_a is exactly $1/\alpha$, i.e., so that all A classes are equiprobable; and with conditional B class probabilities equal to those of the original population. That is to say, it may seem reasonable to replace the quantities ρ_{ab} by the quantities

$$\frac{1}{\alpha}\frac{\rho_{ab}}{\rho_{a\cdot}} \tag{16}$$

and use this as the population to which λ_b is applied. We may thus define, in terms of the conditional probabilities given A_a,

$$\lambda_b{}^* = \frac{\dfrac{1}{\alpha}\sum_a \dfrac{\rho_{am}}{\rho_{a\cdot}} - \dfrac{1}{\alpha}\operatorname{Max}_b \sum_a \dfrac{\rho_{ab}}{\rho_{a\cdot}}}{1 - \dfrac{1}{\alpha}\operatorname{Max}_b \sum_a \dfrac{\rho_{ab}}{\rho_{a\cdot}}}. \tag{17}$$

If we do this in the present example, we get, of course, the same altered ρ table for both hospitals

.477	.023	.500
.125	.375	.500
.602	.398	1.00

and in both cases

$$\lambda_b{}^* = \frac{.250}{.398} = .628.$$

An analogous procedure could be used to define $\lambda_a{}^*$ and λ^*. Note also that other 'artificial' marginal ρ's besides .5 could be used if appropriate. Yule [15] suggests as a desideratum for coefficients of association their invariance under transformations on the $\{\rho_{ab}\}$ matrix of form

$$\rho_{ab} \rightarrow s_a t_b \rho_{ab},\ s_a,\ t_b > 0; \quad a = 1, \cdots, \alpha; \quad b = 1, \cdots, \beta.$$

Such a transformation may readily be found (at least when no $\rho_{ab}=0$) to make *all* four marginals of a two by two table equal to .5. In this connection, we refer to a recent article by Pompilj [10] in which such transformations are carefully discussed.

All further measures may be considered for unweighted or weighted marginal proportions, whichever are appropriate.

6. MEASURES BASED UPON OPTIMAL PREDICTION OF ORDER

6.1. *Preliminaries*

Heretofore we have considered measures of association suitable for the unordered case, that is, measures which do not change if the columns (rows) are permuted. Now we shall suggest a measure suitable for the ordered case. Suppose that the situation is of the following kind:

(*i*) Two polytomies, A and B.
(*ii*) No relevant underlying continua.
(*iii*) Directed ordering is of interest.
(*iv*) The two polytomies appear symmetrically.

By (*iii*) we mean that we wish to distinguish, in the 3×3 case between, for example,

ρ_{11}	0	0
0	ρ_{22}	0
0	0	ρ_{33}

and

0	0	ρ_{13}
0	ρ_{22}	0
ρ_{31}	0	0

calling the first of these complete association and the second complete counterassociation. We may wish to make the convention that in these two cases the proposed measure should take the values $+1$ and -1 respectively. If the sense or direction of order is irrelevant we can, for example, simply take the absolute value of a measure appropriate to directed ordering.

There are vaguenesses in the idea of complete ordered association. For example, everyone would probably agree that the following case is one of complete association:

0	0	0
ρ_{21}	0	0
0	ρ_{32}	0

The following situation is not so clear:

ρ_{11}	0	0
ρ_{21}	ρ_{22}	0
0	ρ_{32}	ρ_{33}

As before, the procedure we shall adopt toward this and toward more complex questions is to base the measure of association on a probabilistic model of activity which often may be appropriate and typical.

6.2. *A Proposed Measure*

Our proposed model will now be described. Suppose that two individuals are taken independently and at random from the population (technically with replacement, but this is unimportant for large populations). Each falls into some (A_a, B_b) cell. Let us say that the first falls in the $(A_{\underline{a}_1}, B_{\underline{b}_1})$ cell, and the second in the $(A_{\underline{a}_2}, B_{\underline{b}_2})$ cell. (Underlined letters denote random variables.) $\underline{a}_i$ $(i=1, 2)$ takes values from 1 to α; $\underline{b}_i$ $(i=1, 2)$ takes values from 1 to β.

If there is independence, one expects that the order of the $\underline{a}$'s has no connection with the order of the $\underline{b}$'s. If there is high association one expects that the order of the $\underline{a}$'s would generally be the same as that of the $\underline{b}$'s. If there is high counterassociation one expects that the orders would generally be different.

Let us therefore ask about the probabilities for like and unlike orders. In order to avoid ambiguity, these probabilities will be taken conditionally on the absence of ties. Set

$$(18)\quad \Pi_s = \Pr\{\underline{a}_1 < \underline{a}_2 \text{ and } \underline{b}_1 < \underline{b}_2; \text{ or } \underline{a}_1 > \underline{a}_2 \text{ and } \underline{b}_1 > \underline{b}_2\}$$

$$(19)\quad \Pi_d = \Pr\{\underline{a}_1 < \underline{a}_2 \text{ and } \underline{b}_1 > \underline{b}_2; \text{ or } \underline{a}_1 > \underline{a}_2 \text{ and } \underline{b}_1 < \underline{b}_2\}$$

$$(20)\quad \Pi_t = \Pr\{\underline{a}_1 = \underline{a}_2 \text{ or } \underline{b}_1 = \underline{b}_2\}.$$

Then the conditional probability of like orders given no ties is $\Pi_s/(1-\Pi_t)$ and the conditional probability of unlike orders given no ties is $\Pi_d/(1-\Pi_t)$. Of course, the sum of these two quantities is one.

A possible measure of association would then be $\Pi_s/(1-\Pi_t)$, but it is a bit more convenient to look at the following quantity:

$$(21)\qquad \gamma = \frac{\Pi_s - \Pi_d}{1 - \Pi_t}$$

or the *difference* between the conditional probabilities of like and unlike orders. In other words γ tells us how much more probable it is to get like than unlike orders in the two classifications, when two individuals are chosen at random from the population.

Since $\Pi_s+\Pi_d=1-\Pi_t$, we may write γ as

$$(22)\qquad \gamma = \frac{2\Pi_s - 1 + \Pi_t}{1 - \Pi_t}$$

which is convenient for computation, using the easily checked relationships

$$(23)\qquad \Pi_s = 2\sum_a \sum_b \rho_{ab}\Big\{\sum_{a'>a}\ \sum_{b'>b} \rho_{a'b'}\Big\}$$

$$(24)\qquad \Pi_t = \sum_a \rho_{a\cdot}^2 + \sum_b \rho_{\cdot b}^2 - \sum_a \sum_b \rho_{ab}^2.$$

Some important properties of γ follow:

(*i*) γ is indeterminate if the population is concentrated in a single row or column of the cross-classification table.

(*ii*) γ is 1 if the population is concentrated in an upper-left to lower-right diagonal of the cross-classification table. γ is -1 if the population is concentrated in a lower-left to upper-right diagonal of the table.

(*iii*) γ is 0 in the case of independence, but the converse need not hold except in the 2×2 case. An example of nonindependence with $\gamma=0$ is

.2	0	.2
0	.2	0
.2	0	.2

For tables up to 5×5 with ρ's expressed to two decimal places computation is fairly rapid. If many tables of the same size are at hand a cardboard template would be convenient. A check on Π_s is to recompute using inverted ordering in both dimensions. γ may be rewritten in terms of the ν's by putting "ν_{ab}" for "ρ_{ab}," etc., and replacing "1" in (22) by "ν^2."

In the 2×2 case we find that

$$\gamma = \frac{\rho_{11}\rho_{22} - \rho_{12}\rho_{21}}{\rho_{11}\rho_{22} + \rho_{12}\rho_{21}}. \tag{25}$$

This is the same as Yule's coefficient of association Q mentioned in Section 4. In this case $\gamma = \pm 1$ if any one cell is empty. For example,

ρ_{11}	0
ρ_{21}	ρ_{22}

gives rise to $\gamma = 1$ always.

Any case of the following forms will give rise to $\gamma = 1$, since a conflict in order is impossible:

ρ_{11}	ρ_{12}	0
0	ρ_{22}	ρ_{23}
0	0	ρ_{33}

ρ_{11}	0	0
ρ_{21}	0	0
ρ_{31}	ρ_{32}	ρ_{33}

The right-hand table might be thought of as a case of "complete curvilinear association."

Stuart [11], starting from a suggestion by Kendall [6], has proposed a measure of association in the ordered case much like γ. Stuart's measure, which he calls τ_c is, in our notation,

$$\tau_c = \frac{\Pi_s - \Pi_d}{(m-1)/m}$$

where $m = \text{Min}(\alpha, \beta)$. The term $(m-1)/m$ is introduced in order that τ_c may attain, or nearly attain, the absolute value 1 when the entire population lies in a longest diagonal of the table. Stuart develops his measure by considering a two-way ordered classification table as two rankings of a population, where many ties appear in one or both rankings as two individuals of the population fall in the same column or row or both. Then each ordered pair of individuals is assigned a score with respect to each ranking: 0 if there is a tie, or ± 1 as one or the other is ranked higher. Finally the product-moment correlation coefficient is formally computed with these scores, and the norming factor is introduced.

Thus, our development of γ is seen to give another and more natural interpretation for the numerator of τ_c: it is the probability of like order less the probability of unlike order when two individuals are chosen at random. In addition the form in which τ_c is given above, together with (23) and (24), suggests a computation procedure somewhat different than that of [11].

6.3. *An Example*

Whelpton, Kaiser, and others [17] have investigated in great detail relationships between human fertility and a number of social and psychological characteristics of married couples. The analyses resulting from these investigations are replete with cross-classification tables, together with accompanying verbal explanations and recapitulations. Numerical indexes of association appear to have been used rarely, if at all, in this work.

We wish to examine briefly one of these cross-classification tables as an example of a cross-classification with an order in both classifications. This examination should be construed neither as approval nor criticism of the methodology used in the studies edited by Whelpton and Kaiser, for this would not be appropriate here. (The reader may refer to [18] and [19] for critical reviews.) However, we do feel that the use of summarizing indexes of association in a study of this kind may well be worth while for at least two reasons. One is that the reader finds it very difficult to obtain a bird's-eye view of the extensive numerical material without depending almost wholly on the author's own conclusions. Second, the use of indexes would mitigate the criticism that the author, consciously or not, selects from his numerical data those comparisons that are in line with his a priori beliefs. Needless to say, an index of association is recommended by these arguments only if it has some reasonable interpretation.

The particular table we wish to consider follows, in terms of numbers of married couples. It refers to a rather special, but well defined, population: white Protestant married couples living in Indianapolis, married in 1927, 1928, or 1929, and so on. The data were obtained by stratified sampling, with strata based on numbers of live births. However, for present purposes we do not consider any questions of sampling, response error, specification of population, etc. The table is condensed from a more detailed cross-classification given in [17], vol. 2, pp. 286, 389, and 402. Further, we shall not define the fertility-planning categories that follow, but merely indicate the order.

CROSS-CLASSIFICATION BETWEEN EDUCATIONAL LEVEL OF WIFE AND FERTILITY-PLANNING STATUS OF COUPLE. SOURCE [17], VOL. 2. NUMBERS IN BODY OF TABLE ARE FREQUENCIES

Highest level of formal education of wife	Fertility-planning status of couple				
	A Most effective planning of number and spacing of children	B	C	D Least effective planning of children	Row totals
one year college or more	102	35	68	34	239
3 or 4 years high school	191	80	215	122	608
less than 3 years high school	110	90	168	223	591
Column totals	403	205	451	379	1438

This is clearly a case where there is relevant order in both classifications. We may first compute Π_s as follows (schematically):

$$\begin{aligned}\Pi_s &= \frac{2}{(1438)^2}[102(80+90+215+168+122+223) \\ &\quad + 35(215+168+122+223)+\cdots+215\,(223)] \\ &= \frac{2}{(1438)^2}[102\times 898+35\times 728+\cdots+215\times 223] \\ &= \frac{2\times 311{,}632}{2{,}067{,}844} = .301.\end{aligned}$$

This means that if we pick two couples at random from those included in the table, the probability is .301 that they are not tied in either classification and that they fall in the same order for both classifications (e.g., if educational level of wife is greater for first couple chosen, then effectiveness of fertility planning is also greater).

Similarly we compute that $\Pi_d = .163$. This is the probability of no ties and *different* orders. Finally Π_t, the probability of a tie in at least one classification, is .536. Note that $\Pi_s+\Pi_d+\Pi_t = 1.000$.

The conditional probability of like order, given no tie, is $\Pi_s/(1-\Pi_t) = .301/.464 = .649$; and the conditional probability of unlike order is $.163/.464 = .351$. Clearly there is a greater chance of like order than of unlike order, and this means positive association, if the operational model is a reasonable one. To measure the magnitude of this association we may use γ, which here is equal to

$$\frac{.301 - .163}{.464} = .298.$$

This is the difference between the conditional probabilities of like and unlike order, given no ties.

It might be thought that one should look, not at the actual population above, but at a related population with equal row totals and with the same relative frequencies within each row. That is, we might wish to work with a derived population within which one-third of the wives lie in each education category, but which is otherwise the same. This derived population is readily obtained (in terms of its ρ_{ab}'s) by dividing each frequency in the above table by three times the total in its row. Very minor adjustments were made because of rounding, in order that the over-all sum be 1.000. For the same reason, the row totals are not exactly equal.

CROSS-CLASSIFICATION BETWEEN EDUCATIONAL LEVEL OF WIFE AND FERTILITY-PLANNING STATUS OF COUPLE. DERIVED FROM PRIOR TABLE BY ADJUSTMENT TO MAKE ROW TOTALS EQUAL. NUMBERS IN BODY OF TABLE ARE RELATIVE FREQUENCIES (ρ_{ab}'s).

Highest level of formal education of wife	Fertility-planning status of couple				Row totals
	A Most effective planning of number and spacing of children	B	C	D Least effective planning of children	
one year college or more	.142	.049	.095	.047	.333
3 or 4 years high school	.105	.044	.118	.067	.334
less than 3 years high school	.062	.050	.095	.126	.333
Column totals	.309	.143	.308	.240	1.000

For this table we find $\Pi_s = .325$, $\Pi_d = .170$, $\Pi_t = .505$.

Hence $\Pi_s/(1-\Pi_t) = .657$, $\Pi_d/(1-\Pi_t) = .343$, and $\gamma = .314$. There is no great difference between the original and the adjusted table in regard to association as measured by probabilities of like and unlike order.

Alternatively, one might wish to adjust the tabular entries so that column totals are equal, or one might attempt to adjust the entries so that the row totals are equal and the column entries are equal.

7. THE GENERATION OF MEASURES BY THE INTRODUCTION OF LOSS FUNCTIONS

7.1. *Models Based on Loss Functions*

Instead of obtaining a measure as a natural function of probabilities in the context of a model of predictive behavior, one can more generally employ loss functions. In such a way, one can even artificially generate the conventional measures described in Section 4.

7.2. *Loss Functions and the* λ *Measures*

In the context of Section 5.1 let us suppose that in guessing an individual's B class one incurs a loss $L(b_1, b_2)$, where B_{b_1} is the true B class and B_{b_2} is the guessed one. Consider first guessing B_b given no information. Then a scheme of guessing B_b with probability $p_b(p_b \geqq 0, \sum p_b = 1)$ leads to an average loss of $\sum_{b_1} \sum_{b_2} \rho_{\cdot b_1} p_{b_2} L(b_1, b_2)$. It is easily seen that this average is minimized by guessing that B_{b_2} for which $\sum_b \rho_{\cdot b} L(b, b_2)$ is a minimum, or if there are two or more minima by guessing any one of them. Let b_L be any one of these b_2's, so that the minimum average loss is $\sum_b \rho_{\cdot b} L(b, b_L)$.

On the other hand if the individual's A class is known to be A_a, the best scheme of guessing is to select b_2 to minimize $\sum_b \rho_{ab} L(b, b_2)$. Let b_{La} be such a minimizing b_2; then the minimum average loss when A_a is known is $\sum_b (\rho_{ab}/\rho_{a\cdot}) L(b, b_{La})$, and the over-all minimum average loss with A_a's known is $\sum_a \sum_b \rho_{ab} L(b, b_{La})$.

The decrease in loss as we pass from the first case to the second is therefore

$$\sum_b \rho_{\cdot b} L(b, b_L) - \sum_a \sum_b \rho_{ab} L(b, b_{La}). \tag{26}$$

It would be reasonable to norm this by division by the first term, $\sum_b \rho_{\cdot b} L(b, b_L)$, to obtain a generalization of λ_b.

Notice that if $L(b_1, b_2)$ is 0 when $b_1 = b_2$ and 1 when $b_1 \neq b_2$, we obtain exactly λ_b. Analogous procedures give us generalizations of λ_a and λ. A slight extension of the procedure, permitting the loss to depend on the true A class as well as the true and guessed B classes, gives a generalization of λ_b^*.

7.3. *The Conventional Measures in Terms of Loss Functions*

Suppose, instead of predicting the classes of individuals, we are asked to determine the values ρ_{ab} when only the $\rho_{a\cdot}$ and $\rho_{\cdot b}$ are known. In the case of independence, these ρ_{ab} are $\rho_{a\cdot}\, \rho_{\cdot b}$. In the more general case, the difference between ρ_{ab} and $\rho_{a\cdot}\, \rho_{\cdot b}$ may be thought of as the amount of error made by assuming independence, If the loss is proportional to the square of the error, inversely proportional to the estimate $\rho_{a\cdot}\, \rho_{\cdot b}$, and additive, we have

$$\sum_a \sum_b k_{ab} \frac{(\rho_{ab} - \rho_{a\cdot}\rho_{\cdot b})^2}{\rho_{a\cdot}\rho_{\cdot b}} \tag{27}$$

where the k_{ab}'s are given constants. For comparison with standard chi-square, express this in terms of the ν_{ab}'s

$$\sum_a \sum_b k_{ab} \frac{\left(\nu_{ab} - \frac{\nu_{a\cdot}\nu_{\cdot b}}{\nu}\right)^2}{\nu_{a\cdot}\nu_{\cdot b}} \tag{28}$$

and finally set $k_{ab}=\nu$ to obtain just the chi-square statistic.

Although this procedure and loss function seem to us rather artificial, they do give one way of motivating the chi-square statistic as a measure of association.

8. RELIABILITY MODELS

8.1. *Generalities*

Consider now cases in which the classes are the *same* for the two polytomies, so that we deal with an $\alpha\times\alpha$ table, but differ in that assignment to class depends on which of two methods of assignment is used. Thus we might for example consider two psychological tests both of which classify deranged individuals as to the type of mental disorder from which they suffer. Or again, we might consider two observers taking part in a sociological experiment wherein they independently and subjectively rate each child in a group of children on a five point scale for degree of cooperation.

One is often concerned in such cases with the degree to which the two methods of assignment to class agree with each other. In the case of the psychological tests, for example, one of the tests might be a well established standard procedure and the other might be a more easily applied variant under consideration as a substitute. The psychologist would probably only consider the variant seriously if it gave the same answers as the standard test often enough in some sense which he would have to explicate. In the case of the two observers, the problem might be whether the kind of subjective ratings given by trained observers in that context are similar enough to warrant the use of such subjective ratings at all.

As before we shall not consider here sampling problems, but rather shall suppose the population ρ_{ab}'s known. The several distinctions and conventions of Sections 2 and 3 apply here of course, but the measures suggested in Sections 5 and 6 do not seem appropriate in this reliability context. One reason is that the classes are the same for both polytomies. This means that even in the unordered case we do *not* want a measure which is invariant under interchange of rows and interchange of columns unless the two interchanges are the same.

An obvious measure of reliability in such a study is just $\sum_a \rho_{aa}$, the probability of agreement. However, we shall also consider some other possibilities.

8.2. *A Measure of Reliability in the Unordered Case*

The measure we shall now propose might be appropriate under the following conditions:

(*i*) Two polytomies are the same, but arise from different methods of assignment to class.
(*ii*) No relevant underlying continua.
(*iii*) No relevant ordering.
(*iv*) Our interest in reliability is symmetrical as between the two polytomies.

A modal class over both classifications is any $A_a(=B_a)$ such that $\rho_{a\cdot}+\rho_{\cdot a}\geqq\rho_{a'\cdot}+\rho_{\cdot a'}$ for all a'. It is simplest to suppose that there is a unique modal class, but if there are more any can be chosen. Denote by $\rho_{M\cdot}$ and $\rho_{\cdot M}$ the two marginal proportions corresponding to the modal class.

A modal class can be given the following interpretation: choose an individual at random from the population and pick one of the two methods of assignment by flipping a fair coin. What is the long-run best guess beforehand of how the chosen method will classify the chosen individual? The answer is: the modal class; and if the modal class is A_a, then the probability of a correct guess is $\frac{1}{2}(\rho_{a\cdot}+\rho_{\cdot a})=\frac{1}{2}(\rho_{M\cdot}+\rho_{\cdot M})$.

In so far as there is good reliability between the two methods of assignment, one could make a better guess if one knew how the other method of assignment would classify the individual, and then followed the rule of guessing the *same* class for the method being predicted. The probability of a correct guess would then be $\sum\rho_{aa}$. Thus as we go from the no information situation to the other-method-known situation, the probability of error decreases by $\sum\rho_{aa}-\frac{1}{2}(\rho_{M\cdot}+\rho_{\cdot M})$. This quantity may vary from $-\frac{1}{2}$ to $1-(1/\alpha)$. It takes the value $-\frac{1}{2}$ when all the diagonal ρ_{aa}'s are zero and the modal probability, $\rho_{M\cdot}+\rho_{\cdot M}$ is 1. It takes the value $1-(1/\alpha)$ when the two methods always agree and each category is equi-probable.

To get a measure we should alter the above quantity, since a sufficiently large ρ_{aa} for some a will make the above quantity low even though $\sum \rho_{aa}$ is nearly 1. It seems reasonable to norm by division by the probability of error given no information, that is by $1-\frac{1}{2}(\rho_{M\cdot}+\rho_{\cdot M})$. Hence we propose the measure

$$\lambda_r = \frac{\sum \rho_{aa} - \frac{1}{2}(\rho_{M\cdot} + \rho_{\cdot M})}{1 - \frac{1}{2}(\rho_{M\cdot} + \rho_{\cdot M})} \,. \tag{29}$$

This may be interpreted as the relative decrease in error probability as we go from the no information situation to the other-method-known situation.

The measure λ_r can take values from -1 to 1. It takes the value -1 when all the diagonal ρ_{aa}'s are zero and the modal probability, $\rho_{M\cdot}+\rho_{\cdot M}$ is 1. It takes the value 1 when the two methods always agree. λ_r is indeterminate only when both methods always give only one and the same class. In the case of independence λ_r assumes no particular value. This characteristic might be considered a disadvantage, but it seems to us that an index of this kind would only be used where there is known to be dependence between the methods, so that misbehavior of the index for independence is not important.

8.3. *Reliability in the Ordered Case*

For the case in which the classes are ordered, but a meaningful metric is absent, we have been unable to find a measure better than one of the following kind:

$$\sum_{a=1}^{\alpha} \rho_{aa} \quad \text{(as suggested in Section 8.1)} \tag{30a}$$

$$\sum_{|a-b|\leqq 1} \rho_{ab} \tag{30b}$$

$$\sum_{|a-b|\leqq 2} \rho_{ab}, \tag{30c}$$

that is, the only reasonable measures we know of are those that are based upon either the probability of agreement, the probability of agreement to within one neighboring class, two neighboring classes, and so on. If desired one could weight these probabilities when classification in a neighboring class is not as desirable as in the same class. Thus one might consider something like $\sum \rho_{aa}+\frac{1}{2}\sum_{|a-b|=1}\rho_{ab}$ or its obvious variants. These measures may also be justified easily by loss-function arguments.

9. PROPORTIONAL PREDICTION

Instead of basing a measure of association on optimal prediction one might consider measures based upon a prediction method which reconstructs the population, in a sense to be described. The use of such a measure was suggested to us by W. Allen Wallis. For simplicity, we restrict ourselves to the asymmetric situation of Section 5.1 where λ_b was constructed. Of course one could apply the same approach in other situations.

Our model of activity, as before, is the following: An individual is chosen at random from the population and we are asked to guess his B class either (1) given no information or (2) given his A class.

Optimal guessing will lead to a definite B class in case (1) and to a definite B class for each A class in case (2) (except that in the case of tied $\rho_{\cdot b}$'s or ρ_{ab}'s we have some choice). While such optimal guessing leads to the lowest average frequency of error, the resulting distribution of guessed classes will usually be very different from the original distribution in the population. For some purposes this might be undesirable and one is led to the following model of activity:

Case *1*. Guess B_1 with probability $\rho_{\cdot 1}$, B_2 with probability $\rho_{\cdot 2}$, $\cdots$, B_β with probability $\rho_{\cdot \beta}$.

Case *2*. Guess B_1 with probability $\rho_{a1}/\rho_{a\cdot}$ (the conditional probability of B_1 given A_a), B_2 with probability $\rho_{a2}/\rho_{a\cdot}$, $\cdots$, B_β with probability $\rho_{a\beta}/\rho_{a\cdot}$.

In each case the guessing is to proceed by throwing a β-sided die whose bth side appears with probability $\rho_{\cdot b}$ (case 1) or $\rho_{ab}/\rho_{a\cdot}$ (case 2). This may be accomplished using a table of "random numbers." If we make many such guesses independently it is plain that we shall approximately reconstruct the marginal distribution of the B_b's (case 1) and the joint distribution of the (A_a, B_b)'s (case 2).

The long-run proportion of *correct* predictions in case (1) will be $\sum_{b=1}^{\beta} \rho_{\cdot b}^2$, and in case (2) it will be $\sum_{a=1}^{\alpha} \sum_{b=1}^{\beta} \rho_{ab}^2/\rho_{a\cdot}$. Hence the relative *decrease* in the proportion of incorrect predictions as we go from case (1) to case (2) is

$$(31) \qquad \tau_b = \frac{\sum_a \sum_b \rho_{ab}^2/\rho_{a\cdot} - \sum_b \rho_{\cdot b}^2}{1 - \sum_b \rho_{\cdot b}^2}$$

which can be readily expressed in the chi-square-like form

$$\tau_b = \frac{\sum_a \sum_b \frac{(\rho_{ab} - \rho_{a\cdot}\rho_{\cdot b})^2}{\rho_{a\cdot}}}{1 - \sum_b \rho_{\cdot b}^2}, \tag{32}$$

It is clear that τ_b takes values between 0 and 1; it is 0 if and only if there is independence, and 1 if and only if knowledge of A_a completely determines B_b. Finally τ_b is indeterminate if and only if both independence and determinism simultaneously hold, that is if all $\rho_{\cdot b}$'s but one are zero.

10. ASSOCIATION WITH A PARTICULAR CATEGORY

A group of modifications of many of the preceding measures arises from the observation that there may be little association between the A and B polytomies in general, but if an individual is in a particular A class it may be easy to predict his B class. Suppose, then, that we want the association between A_{a_0}, a specific A class, and the B polytomy. One need only condense all the A_a rows where $a \neq a_0$ into a single row, thus obtaining a $2 \times \beta$ table, and apply whatever measure of association is thought appropriate. The table will have this appearance.

	B_1	B_2	$\cdots$	B_β
A_{a_0}	$\rho_{a_0 1}$	$\rho_{a_0 2}$	$\cdots$	$\rho_{a_0 \beta}$
$A_a \ (a \neq a_0)$	$\rho_{\cdot 1} - \rho_{a_0 1}$	$\rho_{\cdot 2} - \rho_{a_0 2}$	$\cdots$	$\rho_{\cdot \beta} - \rho_{a_0 \beta}$

We are indebted to L. L. Thurstone for pointing out to us the importance of this modification.

11. PARTIAL ASSOCIATION

When there are more than two polytomies it is natural to think of partial association between two of them with the effect of the others averaged out in some sense. Two such measures of partial association will be suggested here for the asymmetrical case and three polytomies. The viewpoint will be that of optimal prediction. Analogous symmetrical measures may be readily obtained, and the restriction to three polytomies is purely for convenience of notation. The first two polytomies will be denoted as before; the third will consist of the classification $C_1, C_2, \cdots, C_\gamma$. The proportion of the population in A_a, B_b, and

C_c is ρ_{abc}, and dots will be used to denote marginal values in the conventional way. The proposed measures will be for partial association between the A and B polytomies 'averaged' over the C polytomy. (Do not confuse the integer γ used here with the index γ of Section 6.)

11.1. *Simple Average of* λ_b

For fixed C_c, we have a conditional $A \times B$ double polytomy with relative frequencies $\rho_{abc}/\rho_{\cdot\cdot c}$. Hence we can compute λ_b for each such table—call it $\lambda_b(c)$ to show dependence on c. Now it might seem natural to average these values with weights equal to the marginal relative frequencies of the C classifications. That is, we suggest

$$\lambda_b(A, B \mid C) = \sum_{c=1}^{\gamma} \rho_{\cdot\cdot c}\lambda_b(c). \tag{33}$$

11.2. *Measure Based Directly on Probabilities of Error*

It seems to us somewhat better, from the viewpoint of interpretation, to proceed as follows. For given C_c if we predict B classes optimally on the basis of no further information, the probability of error is $1-(\mathrm{Max}_b\ \rho_{\cdot bc})/\rho_{\cdot\cdot c}$; whereas if we know the A class the probability of error is $1-(\sum_a \mathrm{Max}_b\ \rho_{abc})/\rho_{\cdot\cdot c}$. Hence, if we are given individuals from the population at random and always told their C class, the probability of error in optimal guessing if we know nothing more is $1-\sum_c \mathrm{Max}_b\ \rho_{\cdot bc}$; whereas if we also know the A class the probability is $1-\sum_c \sum_a \mathrm{Max}_b\ \rho_{abc}$. Thus the relative decrease in probability of error is

$$\lambda_b'(A, B \mid C) = \frac{\sum_c \sum_a \mathrm{Max}_b\ \rho_{abc} - \sum_c \mathrm{Max}_b\ \rho_{\cdot bc}}{1 - \sum_c \mathrm{Max}_b\ \rho_{\cdot bc}} \tag{34}$$

which might often be a satisfactory measure of partial association.

12. MULTIPLE ASSOCIATION

When there are more than two polytomies one may well be interested in the multiple association between one of them and all the others. One simple way of handling this in the unordered case will be described here for three polytomies A, B, and C as defined in Section 11. We suppose that the multiple association between A and B-together-with-C is of interest. Simply form a two-way table whose rows represent the A polytomy and whose columns represent all combinations B_b, C_c and then apply the appropriate two-polytomy measure. The table will have this appearance:

	B_1C_1	B_1C_2	$\cdots$	B_1C_γ	B_2C_1	$\cdots$	B_2C_γ	$\cdots$	$B_\beta C_\gamma$
A_1	ρ_{111}	ρ_{112}	$\cdots$	$\rho_{11\gamma}$	ρ_{121}	$\cdots$	$\rho_{12\gamma}$	$\cdots$	$\rho_{1\beta\gamma}$
A_2	ρ_{211}	ρ_{212}	$\cdots$	$\rho_{21\gamma}$	ρ_{221}	$\cdots$	$\rho_{22\gamma}$	$\cdots$	$\rho_{2\beta\gamma}$
$\vdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$	$\cdots$	$\vdots$
A_α	$\rho_{\alpha 11}$	$\rho_{\alpha 12}$	$\cdots$	$\rho_{\alpha 1\gamma}$	$\rho_{\alpha 21}$	$\cdots$	$\rho_{\alpha 2\gamma}$	$\cdots$	$\rho_{\alpha\beta\gamma}$

Note that this procedure does not take the $B \times C$ association into account. There is a rough analogy here with the motivation for the standard multiple correlation coefficient of normal theory. The standard multiple correlation coefficient may be (and often is) motivated by defining it as the maximum correlation coefficient obtainable between a given variate and linear combinations of the other variates. That is, it is a measure of association between a given variate and the best estimate (in a certain sense) of that variate based upon all the other variates. It is true that the standard multiple correlation coefficient may be expressed as a function of the several ordinary bivariate correlation coefficients, but in a sense this is a consequence of the strong structural assumption of multivariate normality.

13. SAMPLING PROBLEMS

The discussion thus far has been in terms of *known* populations, whereas in practice one generally deals with a sample from an *unknown* population. One then asks, given a formal measure of association, how to estimate its value, how to test hypotheses about it, and so on.

Exact sampling theory for estimators from cross-classification tables is difficult to work with. However, the asymptotic theory is reasonably manageable, at least in some cases. We intend to discuss this in another paper, where we shall state some of the asymptotic distributions and say what we can of their value as approximations.

14. CONCLUDING REMARKS

The aim of this paper has been to argue that measures of association should not be taken blindly from the handiest statistics textbook, but rather should be carefully constructed in a manner appropriate to the problem at hand. To emphasize and illustrate this argument we have described a number of such measures which we feel might be useful in several situations. While we naturally take a friendly view towards these measures, we can hardly claim that they are more than examples.

This methodologically neutral position should not be carried to an extreme. It would be ridiculous to ask each empirical scientist in each separate study to forge afresh new statistical tools. The artist cannot paint many pictures if he must spend most of his time mixing pigments. Our belief is that each scientific area that has use for measures of association should, after appropriate argument and trial,[4] settle down on those measures most useful for its needs.

15. REFERENCES

[1] Cramér, Harald, *Mathematical Methods of Statistics*, Princeton, Princeton University Press (1946)

[2] Fisher, R. A., *Statistical Methods for Research Workers*, Tenth Edition, New York, Hafner Publishing Co. (1948)

[3] Greenwood, Major, Jr., and Yule, G. Udny, "The statistics of anti-typhoid and anti-cholera inoculations, and the interpretation of such statistics in general," *Proceedings of the Royal Society of Medicine*, 8 [part 2] (1915), 113–94.

[4] Guttman, Louis, "An outline of the statistical theory of prediction," Supplementary Study B-1 (pp. 253–318) in Horst, Paul and others (editors), *The Prediction of Personal Adjustment*, Bulletin 48, Social Science Research Council, New York (1941).

[5] Jahn, Julius A., "The measurement of ecological segregation: derivation of an index based on the criterion of reproducibility," *American Sociological Review*, 15 (1950), 101–4.

[6] Kendall, M. G., "Rank and product-moment correlation," *Biometrika*, 36 (1949), 177–93.

[7] Kendall, Maurice G., *The Advanced Theory of Statistics*, London, Charles Griffin and Co., Ltd. (1948).

[8] McCormick, Thomas C., "Toward causal analysis in the prediction of attributes," *American Sociological Review*, 17 (1952), 35–44.

[9] Pearson, Karl, and Heron, David, "On theories of association," *Biometrika*, 9 (1913), 159–315.

[10] Pompilj, G., "Osservazioni sull'omogamia: La trasformazione di Yule e il limite della trasformazione ricorrente di Gini," *Rendiconti di Matematica e*

[4] For examples of such argument and trial in the field of sociology see J. J. Williams [13], Jahn [5], and McCormick [8].

delle sue Applicazioni, Università di Roma, Istituto Nazionale di Alta Matematica, Ser. V, Vol. 9 (1950), 367–88.

[11] Stuart, A., "The estimation and comparison of strengths of association in contingency tables," *Biometrika*, 40 (1953), 105–10.

[12] Williams, E. J., "Use of scores for the analysis of information in contingency tables," *Biometrika*, 39 (1952), 274–89.

[13] Williams, Josephine J., "Another commentary on so-called segregation indices," *American Sociological Review*, 13 (1948), 298–303.

[14] Youden, W. J., "Index for rating diagnostic tests," *Cancer*, 3 (1950), 32–5.

[15] Yule, G. Udny, "On the methods of measuring association between two attributes," *Journal of the Royal Statistical Society*, 75 (1912), 579–642.

[16] Yule, G. Udny, and Kendall, M. G., *An Introduction to the Theory of Statistics*, London, Charles Griffin and Co., Ltd. (1950)

[17] Whelpton, P. K., and Kiser, Clyde V., *Social and Psychological Factors Affecting Fertility*, Milbank Memorial Fund, New York.
   Volume 1 (1946), *The Household Survey in Indianapolis.*
   Volume 2 (1950), *The Intensive Study; Purpose, Scope, Methods, and Partial Results.*
   Volume 3 (1952), *Further Reports on Hypotheses in the Indianapolis Study.*

[18] Jaffe, A. J., review of [17], Volume 2, *Journal of the American Statistical Association*, 47 (1952), 348–9.

[19] Lewis-Faning, E., review of [17], Volume 3, *Journal of the American Statistical Association*, 49 (1954), 190–3.